A MOLECULAR DYNAMICS INVESTIGATION OF THE TIME-DEPENDENT STOKES SHIFT OF TRYPTOPHAN-28 OF LYSOZYME

Katarina Odak

The Ohio State University
March 27, 2020

Undergraduate Research Thesis
Presented in Partial Fulfillment of the Requirements for graduation
with Honors Research Distinction in Chemistry
in the undergraduate colleges of
The Ohio State University

Advisor: Sherwin J. Singer

Abstract

The time-dependent fluorescence of tryptophan, an amino acid residue with an indole chromophore, reports on the dynamics of the chromophore's environment as it relaxes in response to the charge redistribution accompanying photo-excitation of the chromophore. As such, the time-dependent fluorescence can provide powerful insights into biomolecular dynamics. However, the capacity to report on environmental dynamics rests on the ability to interpret the fluorescence signal on a molecular level. Here, detailed atomistic simulation can provide understanding of the photophysical behavior underlying the experimental signal. In simulations, the time-dependent fluorescence Stokes shift, $S(t)$, is calculated from the change in potential energy of the chromophore's surroundings, consisting of the rest of the protein, water, and ions, as the components respond to the change in the chromophore's charge distribution. Deconstructing the Stokes shift into three regions, there is general agreement on the dynamics of the inertial decay in the sub-picosecond region, as well as the water librations in the 1-2 picosecond region. However, disagreement arises pertaining to the dynamics of the long-timescale region, which ranges from tens to thousands of picoseconds depending on the protein. Proponents of the "biological water" model, which explains that there is an rigid hydration layer surrounding a protein, suggest that long-timescale dynamics are an intrinsic feature of water. Conversely, we, as well as others, advance that the slow dynamics result from protein motion as the system relaxes in response to photo-excitation of the chromophore.

Because simulations of tryptophan-28 of lysozyme have been proposed as evidence of "biological water", the primary motivation of this work is to simulate the time-dependent fluorescence of Trp-28, provide detailed analysis of the origin of the long-timescale dynamics, and, ultimately, debunk the claim that the long-timescale dynamics arise from intrinsically slow water. Using non-equilibrium molecular dynamics, we simulate the systematic response to the charge redistribution following photo-excitation in a flexible system in which all species are capable of responding, a frozen protein system, and a frozen protein and ions system. We find that the water,

protein, and ionic contributions to the Stokes shift have slow components in the flexible system. With only protein frozen, we find that the long-timescale component of the water contribution remains and is nearly, if not exactly, equal in magnitude and opposite to the ionic contribution. Furthermore, the long-timescale components of all contributions, including that of the total Stokes shift, disappear when both protein and ion motions are frozen. Addressing our secondary motivation of this work, we correct a dielectric continuum model suggesting that protein and water should have equal and opposite responses. While the dielectric continuum model is not valid as originally proposed, it does apply in a different context and explains our observation that the ions and water have equal and opposite long-timescale responses resulting from slow diffusion of ions and coupled ion-water motion from the water molecules in the polarization layer of counter-charge surrounding the ions. Therefore, we conclude that the long-timescale dynamics in the water contribution to the Stokes shift of tryptophan-28 of lysozyme arise from coupled protein-water and ion-water motions, not intrinsically slow water in the "biological water" model.

Acknowledgements

*"If we want to abandon the use of the term **"biological water"**, there is no other way to convey how unique and important water is for biomolecules and, of course, **for life!**"*

– Bagchi et. al., 2017

*"It seems fair to say that the original idea discussed at the beginning of this subsection of a **"biological water"** scenario– an intrinsically slow ice-like hydration shell moving rigidly with a protein– has now been **completely discounted**."*

– Laage, Elsaesser, and Hynes, 2017

These quotations were shown to me by my mentor, Dr. Sherwin J. Singer, on my first day of research during my freshman year of my undergraduate career, with disclaimers like, "Look at these juicy quotes! See how controversial this research is in the realm of photophysics!" Though I joked about the seriousness of the work then, the last three years of studying this topic have made this work very important to me. Before finishing this work, it took me through two graduate courses in chemical physics, a multi-year study of computer programming, and an independent study in mathematics that feels particularly triumphant because I always whined and complained the most about the "math part", but understanding the mathematical theory in both the analysis and computer simulations I conducted is the part of this project of which I am now most proud.

I cannot thank Dr. Singer enough for spending the last three years mentoring me. He seems to have always known what was best for me and what I needed to do to reach my goals, even when I didn't. He always pushed me to pursue that which seemed to be almost too challenging for me. Each time, it wasn't. Each time, I became a better scientist, as well as a better thinker. I worked tirelessly on these challenges, but I will never forget the hand that was guiding me through them.

I would also like to acknowledge Dr. Terry Gustafson. Though he didn't necessarily help with the science of this project, he certainly played an important role in

this work by helping me become the researcher who developed it: a person who values the scientific process, the power of knowledge, and the importance of analytical thinking as part of the lens through which I view life.

Lastly, I would like to thank my team outside of academia. To Deda, thank you for always believing in "super excellence". To my mother and father, thank you for raising me to be an individual who enjoys learning, thinking, and optimizing. To my sisters, thank you for all the years of honest praise and criticism that helped set the standards to which I hold myself. To my partner, AJ, thank you for being just that through my undergraduate career– a true partner. You keep the boat steady whenever I get a little lost in the waves.

In the end, this work will be one of my life's sweetest memories. To me, this work represents my first "thought" that I have to offer the scientific community. I choose to begin with the "juicy quotes" because they originate in dedication to the subject, and over the last few years, this subject has become very important to me, too. Perhaps, one day, I'll even have my own "juicy" comment about the long-timescale dynamics of the fluorescence Stokes shift in a chemical journal.

Contents

List of Figures

Chapter 1

Introduction and Background

1.1 Overview of Tryptophan Fluorescence

Fluorescence occurs when a molecule, previously promoted to an excited state by the absorption of a photon, subsequently emits light, generally at a different wavelength than the incident radiation.[1] The fluorescence of a folded protein is a combination of the fluorescence of individual amino acid residues, including that of tryptophan. Of the three fluorescent amino acid residues in proteins, tryptophan is the most abundant with the highest quantum yield in the ultraviolet range.[2] Additionally, the fluorescence of the indole chromophore of tryptophan demonstrates high sensitivity to fluctuations of its environment, rendering it an excellent molecular probe for monitoring protein structural changes and the hydration dynamics of the protein's surrounding environment.

A non-invasive and highly sensitive technique, fluorescence spectroscopy has the potential to be an excellent tool for studying biological events, including use in constructing protein energy landscapes as well as providing a method of studying enzyme activity.[3] In tryptophan fluorescence, the environment, consisting of water, the protein, and the ions in solution, relaxes in response to the charge redistribution following excitation of the indole chromophore of the tryptophan residue of the protein. This systematic relaxation is highly representative of protein hydration dynamics which are fundamental to protein structure and function.[4–8] Furthermore, tryptophan fluorescence is particularly informative due to its status as an intrinsic chromophore in proteins, eliminating the need for tedious labeling and the possible introduction of artifacts that would arise when using an external probe.[1] This work seeks to elucidate the photophysical nature of tryptophan fluorescence to further its utility as a molecular probe of protein dynamics.

1.2 Time-Dependent Fluorescence Stokes Shift

According to the Franck-Condon principle, the charge distribution within a chromophore shifts following the moment of photo-excitation without significant change in the position of the nuclei. Before excitation, the nuclei tend to be located in the lowest free energy configurations of the ground state, which are generally different from the lowest free energy configurations of the excited state. This results in an initial excited state configuration that is energetically unfavorable. As a result, the surroundings adapt to the excited state charge distribution of the chromophore. Thus, when the chromophore fluoresces, the frequencies of the emitted light generally decrease over time, as shown in Figure 1.1.

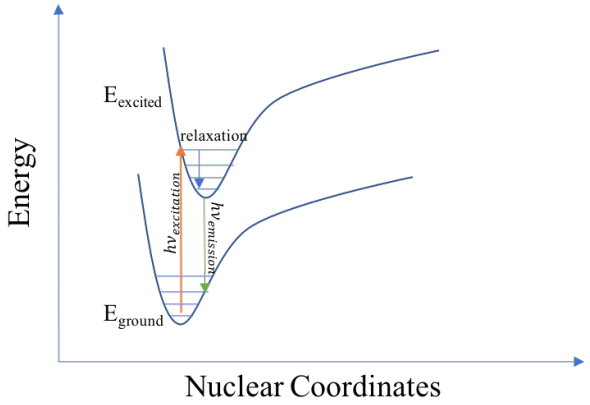

Figure 1.1: Following to the Franck-Condon principle, the nuclei are stationary upon absorption of a photon. The molecule will then relax to reach a lowest free energy configuration of the excited state before fluorescing.

The difference between the emission and excitation photon energy as a function of time gives the time-dependent fluorescence Stokes shift (TDFSS), as shown in Equation (1.1).

$$S(t) = h\nu_{emission}(t) - h\nu_{excitation}(t) \tag{1.1}$$

For the case of a tryptophan residue with an indole chromophore (Figure 2.1), the change in photon energy with respect to time can be modeled in simulations by tracking the change in potential energy of the surroundings (protein, water, and ions) as they relax in response to a change in the chromophore's charge distribution. Thus, the indole chromophore of a tryptophan residue is a molecular probe of protein hydration dynamics.

The TDFSS is often segmented into three pieces, as shown in Figure 1.2. The initial inertial decay occurs rapidly, governed by the inertial motion of the solvent molecules.[9,10] We begin the discussion of the inertial component by modeling the Stokes shift explicitly as the difference between the ground and excited state energies

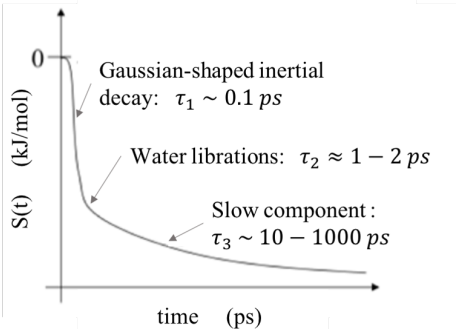

Figure 1.2: The Stokes shift

at time t minus that at time $t = 0$, as shown in Equation (1.2).

$$S(t) = \langle [V_e(t) - V_g(t)] - [V_e(0) - V_g(0)] \rangle_g \tag{1.2}$$

For simplicity of notation, the dependence of the ground and excited state potentials on atomic coordinates, $V_e(t) = V_e(\mathbf{r_1}(t), \mathbf{r_2}(t)...\mathbf{r_N}(t))$, is not shown. The angle brackets indicate averaging over initial conditions, a thermal distribution in the ground state. The energy gap, $V_e(t) - V_g(t)$, changes with time due to the time-evolution of atomic position, $r_{i,\alpha}$ ($i = 1, ..., N$, where N is the number of atoms, and $\alpha = x, y, z$). Because the inertial drop concerns short time behavior, $r_{i,\alpha}(t)$ is expanded in a Taylor series about $t = 0$, as shown in Equation (1.3). In this equation, repeated indices are summed.

$$\begin{aligned} V_e(t) =& V_e(0) + \frac{\partial V_e}{\partial r_{i\alpha}}(\dot{r}_{i\alpha}(0)t + \frac{1}{2m_i}F_{e,i\alpha}(0)t^2 + ...) \\ &+ \frac{1}{2}\frac{\partial V_e}{\partial r_{i\alpha}\partial r_{j\beta}}(\dot{r}_{i\alpha}(0)t + \frac{1}{2m_i}F_{e,i\alpha}(0)t^2 + ...)(\dot{r}_{j\beta}(0)t + \frac{1}{2m_j}F_{e,j\beta}(0)t^2 + ...) \end{aligned} \tag{1.3}$$

In Equation (1.3), Newton's equation is used to express $\ddot{r}_{i\alpha}(0) = \frac{1}{m_i}F_{e,i\alpha}(0)$, where $F_{e,i\alpha} = -\frac{\partial V_e}{\partial r_{i\alpha}}$. Excited state forces are used because relaxation occurs in the excited state. Averaging over a ground state canonical distribution of initial conditions for $[r_i, \dot{r}_i]$ and using $\langle \dot{r}_{i\alpha} \rangle = 0$, Equation (1.3) becomes Equation (1.4).

$$\langle V_e(t) \rangle_g = \langle V_e(0) \rangle_g - \frac{1}{2m_i}\left\langle \frac{\partial V_e}{\partial r_{i\alpha}}\left(\frac{\partial V_e}{\partial r_{i\alpha}} - \frac{\partial V_g}{\partial r_{i\alpha}}\right)\right\rangle_g t^2 + \mathcal{O}(t^3) \tag{1.4}$$

After a similar procedure generates the ground state potential energy at a given time, $V_e(t)$, Equation (1.2) becomes Equation (1.5), showing the initial Gaussian-shaped

inertial drop.

$$S(t) = -\frac{1}{2m_i}\left\langle \left(\frac{\partial V_e}{\partial r_{i\alpha}} - \frac{\partial V_g}{\partial r_{i\alpha}}\right)^2 \right\rangle_g t^2 + \mathcal{O}(t^4) \tag{1.5}$$

Equation (1.5) indicates that the inertial drop depends on the interactions that change between the ground and excited states. Therefore, if water-water interactions are the same in the ground and excited states, as they are in our model, the calculated inertial drop is independent of the choice of water interaction model.

The timescale associated with inertial drop is generally agreed to be in the sub-picosecond range, $\tau_1 \sim 0.1$ picoseconds. Additionally, there is a second short-timescale component of the Stokes shift with time constant $\tau_2 \approx 1 - 2$ picoseconds. These dynamics are understood to arise from hindered rotations, or "librations," of the solvent.[11,12]

Unlike the first two components of the Stokes shift, the interpretation of the long-timescale third component is controversial, with time constant τ_3 ranging from a few tens of picoseconds to the nanosecond range, depending on the protein. The slow component of the TDFSS has been simulated in previous works using both non-equilibrium simulations and linear response theory. This work focuses on using non-equilibrium simulations which best reflect the nature of the TDFSS as a non-equilibrium process.

In a series of non-equilibrium molecular dynamics simulations, initial configurations sampled from a ground-state Boltzmann distribution were propagated from time $t = 0$ on the excited-state potential energy surface, simulating relaxation following the moment of photoexcitation. The total Stokes shift as a function of time, $S_{total}(t)$, was then estimated to be the average over initial conditions of the time-dependent energy difference between the excited and ground states, $\Delta E_{total}(t)$, as shown in Equation (1.6).

$$S_{total}(t) = \langle \Delta E_{total}(t) \rangle_g - \langle \Delta E_{total}(0) \rangle_g \tag{1.6}$$

After normalization, Equation (1.6) becomes Equation (1.7).

$$S'_{total}(t) = \frac{\langle \Delta E_{total}(t) \rangle_g - \langle \Delta E_{total}(\infty) \rangle_g}{\langle \Delta E_{total}(0) \rangle_g - \langle \Delta E_{total}(\infty) \rangle_g} \tag{1.7}$$

$S_{total}(t)$ in Equation (1.6) serves as a measure of the system relaxation occurring as it tends toward equilibrium on the excited-state potential energy surface.

$S_{total}(t)$ depends the sum of the interactions between the chromophore and the different components of its environment. In our model, the potential energy as

a function of atomic position, $V_{total}(\mathbf{r})$, is a sum of Coulomb and Lennard-Jones potentials, as shown in Equation (1.8).

$$\begin{aligned} V_{total}(\mathbf{r}) &= V_{Coulomb}(\mathbf{r}) + V_{LJ}(\mathbf{r}) \\ &= \sum_{i \in indole} \sum_{j \in all} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i \in indole} \sum_{j \in all} 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \end{aligned} \tag{1.8}$$

As the van der Waals potential between the ground and excited states is the same, the difference between the two states rests on the Coulombic interactions between the indole chromophore and its environment, as shown in Equation (1.9).

$$\Delta E_{total}(t) = \sum_{i \in indole} \sum_{j \in all} [V_{excited}(\mathbf{r}_i - \mathbf{r}_j) - V_{ground}(\mathbf{r}_i - \mathbf{r}_j)] \tag{1.9}$$

With the environment consisting of the other atoms in the indole chromophore, the rest of the protein (excluding the chromophore), the water molecules, and the ions, Equation (1.9) becomes Equation (1.10). The indole-indole term is nearly constant, so it is excluded from our Stokes shift in Equation (1.10).

$$\begin{aligned} \Delta E_{total}(t) = & \sum_{i \in indole} \sum_{j \in protein} [V_{excited}(\mathbf{r}_i - \mathbf{r}_j) - V_{ground}(\mathbf{r}_i - \mathbf{r}_j)] \\ & + \sum_{i \in indole} \sum_{j \in water} [V_{excited}(\mathbf{r}_i - \mathbf{r}_j) - V_{ground}(\mathbf{r}_i - \mathbf{r}_j)] \\ & + \sum_{i \in indole} \sum_{j \in ions} [V_{excited}(\mathbf{r}_i - \mathbf{r}_j) - V_{ground}(\mathbf{r}_i - \mathbf{r}_j)] \\ = & \Delta E_{protein}(t) + \Delta E_{water}(t) + \Delta E_{ions}(t) \end{aligned} \tag{1.10}$$

After subtracting $\Delta E(0)$ and averaging over initial conditions, we calculate $S_{total}(t)$ using Equation (1.11).

$$S_{total}(t) = S_{protein}(t) + S_{water}(t) + S_{ions}(t) \tag{1.11}$$

Alternatively, linear response theory has been used to model the Stokes shift. Linear response theory treats the change in the charge distribution of the chromophore as a perturbation and estimates the subsequent relaxation of the system to linear order in the perturbation. Using the fluctuation-dissipation theorem, $\langle \Delta E_{total}(t) \rangle_g$ is treated as a non-equilibrium property with fluctuations in the equilibrated system described by Equation (1.12).

$$\delta \Delta E_{total}(t) = \Delta E_{total}(t) - \langle \Delta E_{total}(t) \rangle_g \tag{1.12}$$

The correlation function $C_{total}(t)$ in Equation (1.13) then corresponds to $S_{total}(t)$ in Equation (1.6).[13,14]

$$C_{total}(t) = \frac{1}{k_B T}\left(\langle \delta\Delta E_{total}(t)\delta\Delta E_{total}(0)\rangle_g - \langle \delta\Delta E_{total}(0)^2\rangle_g\right) \tag{1.13}$$

Using partial correlation functions,[14] the normalized correlation function $C'_{total}(t)$ is generated in Equation (1.14), corresponding to the normalized Stokes shift, $S'(t)$, in Equation (1.7).

$$\begin{aligned} C'_{total}(t) &= \frac{\langle \delta\Delta E_{total}(t)\delta\Delta E_{total}(0)\rangle_g}{\langle \delta\Delta E_{total}(0)^2\rangle_g} \\ &= \frac{\langle \delta\Delta E_{protein}(t)\delta\Delta E_{total}(0)\rangle_g}{\langle \delta\Delta E_{total}(0)^2\rangle_g} + \frac{\langle \delta\Delta E_{water}(t)\delta\Delta E_{total}(0)\rangle_g}{\langle \delta\Delta E_{total}(0)^2\rangle_g} + \frac{\langle \delta\Delta E_{ions}(t)\delta\Delta E_{total}(0)\rangle_g}{\langle \delta\Delta E_{total}(0)^2\rangle_g} \\ &= C'_{protein}(t) + C'_{water}(t) + C'_{ions}(t) \end{aligned} \tag{1.14}$$

The origin of the long-timescale dynamics, modeled numerically in simulations through the methods described above, motivates this work. One theory explains that the slow dynamics are an inherent feature of water surrounding the protein, while another, supported in this work, explains that it is the protein's motion in aqueous environment as it relaxes in response to a changing charge distribution in the chromophore that results in the long-timescale behavior.

1.3 The Biological Water Model

The notion that a rigid layer of water encapsulates proteins and other biomolecules originated in an interpretation of experimental studies which suggested that a protein rotates in solution as if the true radius was larger than the crystallographic radius.[15] This observation was expanded into the idea that the properties of water molecules in the vicinity of proteins significantly differ from water molecules farther away, and, in a way, distinct from water near non-biological surfaces.[15,16] The "biological water" model proposes that a rigid aqueous shell, often referred to as a hydration layer, surrounding a protein consists of "bound" water molecules near the protein that diffuse much more slowly than the "free" water molecules that constitute bulk water.[17,18] The designation as "free" or "bound" is related to timescales as "free" molecules diffuse and reorient, while "bound" molecules can adjust their configuration only by exchange with "free" water, as shown in Figure 1.3.[17–19]

The slow component of the Stokes shift has been used as evidence for biological

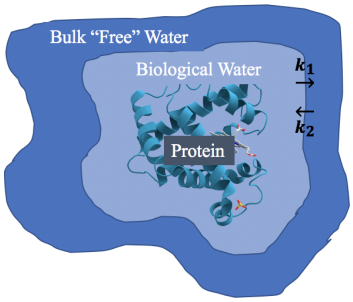

Figure 1.3: Biological Water: "Bound" biological water within close proximity to the protein is rapidly exchanged with "free" bulk water, with the exchange rate consisting of k_1 and k_2. This figure is adapted from work by Bagchi and Zewail et al.[17–21]

water. In an early and influential work, the hydration correlation function, $C'(t)$, obtained from experimental data of tryptophan fluorescence was fit to an additive exponential decay: τ_1 in the sub-picosecond range, τ_2 in the 1-2 picosecond range τ_3 in the 20–50 picosecond range.[18] The fast components with time constants, τ_1 and τ_2, involve the inertial decay and the water librations, respectively.[18,22] The other time constant, $\tau_3 \sim 20 - 50$ picoseconds, has been interpreted as evidence for the presence of a protein hydration layer, with a long-timescale component of the water contribution to the Stokes shift, $\tau_{3,water}$, that is significantly different from that of bulk water, where all relaxation near a single tryptophan residue is over after a few picoseconds.[18,19,23]

1.4 Slow Dynamics Due to Protein: An Alternative to Biological Water

The slow component of the time-dependent fluorescence Stokes shift has also been attributed to the dynamics of the protein itself.[14,24] This theory suggests that the previously described τ_3 component of the Stokes shift is not supportive of the biological water model, but rather that the long-timescale component of the TDFSS varies from $\sim$ 10 picoseconds to $\sim$ 1 nanosecond depending on the protein-related features of the environment. Protein-related features, as described here, are characteristics of the chromophore's local environment that are attributed to the motion of the rest of the protein. The TDFSS has even been shown to vary based on the position of the chromophore probe with respect to the slowly moving biomolecule.[24–27] The long-timescale component after a few picoseconds must then reflect protein dynamics or, at least, strongly coupled protein-water motions.[14,27–31] In this view, the long-timescale component of the Stokes shift is not evidence for the existence of a

slow hydration layer around the protein as part of the "biological water" model.[32,33]

In the years 2006-2008, Hassanali, Li, and Singer proposed a simulation technique to identify the origin of the slow component of the Stokes shift.[27–29] In so-called frozen protein simulations, the protein is immobilized at the instant of photo-excitation and thus unable to relax in response to changing charge distribution of the chromophore. If the biological water model was valid, the slow τ_3 component should still be present with protein frozen. Alternatively, if slow dynamics arise from protein motion, the τ_3 component should disappear with protein frozen. Our group performed calculations for myoglobin, staph nuclease, monellin, and thioredoxin. In every case, the τ_3 component disappeared with protein frozen.[27–29] Of course, there can be no protein motion without water displacement, and Li, Hassanali, Kao, Zhong, and Singer described the slow component as coupled protein-water motion.

Seeking to bolster the "biological water" model, Bagchi and co-workers recently claimed that slow dynamics persisted in the Stokes shift of tryptophan-28 of lysozyme after freezing protein motion, as shown in Figure 1.4.[23] Therefore, our primary aim in this work is to use molecular dynamics simulations of the relaxation of lysozyme following the photo-excitation of the indole chromophore of the tryptophan-28 residue to test the validity of Bagchi's claim.

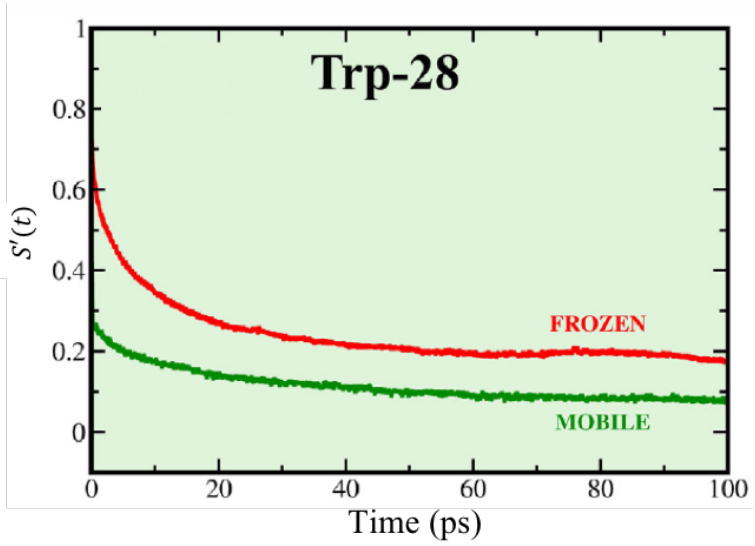

Figure 1.4: The normalized Stokes shift of Trp-28 from Bagchi et al. 2017, who claim that a slow component remains in the Stokes shift when the protein is frozen

Halle and Nilsson suggest that there is a slow water component to the Stokes shift that is coupled, but not dynamically, to a slow protein component.[14,31] They suggest that water and protein responses are approximately equal in magnitude and opposite, which implies that it is erroneous to identify the slow component of the Stokes shift as coupled water-protein fluctuations.[27–29] However, we discount this

claim because it uses a dielectric continuum model for water as an aqueous solvent, and molecular dynamics simulations do not agree with its predictions.

In the Halle and Nilsson picture, the total difference between the ground and excited state interaction energies, taken as a a function of protein conformation, $\Delta E_{total}(x_p)$, where x_p stands for the coordinates of the protein, is assumed to be dominated by Coulomb interactions. In water, modeled as a continuum with dielectric constant ϵ, $\Delta E_{total}(x_p)$ can be calculated from the gas phase chromophore-protein interaction potential, $\Delta E^o_{protein}(x_p)$, as shown in Equation (1.15).

$$\Delta E_{total}(x_p) = \frac{1}{\epsilon}\Delta E^o_{protein}(x_p) \tag{1.15}$$

As the system consists only of the indole, the rest of the protein, and the continuum solvent, the water contribution to $\Delta E_{total}(x_p)$ can be calculated as the total difference minus the protein contribution, as shown in Equation (1.16).

$$\begin{aligned}\Delta E_{water}(x_p) &= \Delta E_{total}(x_p) - \Delta E_{protein}(x_p) \\ &= \frac{1}{\epsilon}\Delta E^o_{protein}(x_p) - \Delta E^o_{protein}(x_p) \\ &= -\left(1 - \frac{1}{\epsilon}\right)\Delta E^o_{protein}(x_p)\end{aligned} \tag{1.16}$$

With water as solvent, $\epsilon \gg 1$, and the water contribution to the interaction between chromophore and surroundings should be approximately equal and opposite to protein contribution, as shown in Equation (1.17).

$$\Delta E_{water}(x_p) \approx -\Delta E^o_{protein}(x_p) \tag{1.17}$$

The Stokes shift, however, is the *difference* between $\Delta E_{total}(x_p)$ at a given time t and the same quantity at time $t = 0$, as shown in Equation (1.18).

$$S_{total}(t) = \Delta E_{total}\left(x_p(t)\right) - \Delta E_{total}(x_p(0)) \tag{1.18}$$

The group-specific contributions to the total Stokes shift are thus represented by Equation Set (1.19).

$$\begin{aligned}S_{protein}(t) &= \Delta E^o_{protein}(x_p(t)) - \Delta E^o_{protein}(x_p(0)) \\ S_{water}(t) &= \Delta E_{water}(x_p(t)) - \Delta E_{water}(x_p(0))\end{aligned} \tag{1.19}$$

Substituting the relationship derived in Equation (1.16), the water contribution to

the Stokes shift can be represented by Equation (1.20).

$$S_{water}(t) = -\left(1 - \frac{1}{\epsilon}\right) S_{protein}(t) \tag{1.20}$$

Equation (1.20) implies that the slow component of the Stokes shift disappears when protein coordinates are fixed at the moment of photo-excitation. Previous studies[29] in which the water contribution disappears when the protein motion is fixed were used as an example of the validity of the dielectric continuum model.[14,31] However, the further prediction from Equation (1.20) that water and protein contributions are nearly equal in magnitude and opposite is not supported by molecular dynamics studies.[27,29] Therefore, testing Halle and Nilsson's claim is a secondary motivation for our work.

Chapter 2

Methodology

2.1 Generating the Stokes Shift

Non-equilibrium molecular dynamics simulations were used to construct the Stokes shift. GROMACS[34] was used to simulate a 100 nanosecond (ns) trajectory of lysozyme[35] in SPC/E water[36] with 8 chlorine (Cl^-) counter-ions added to neutralize the protein's positive charge. This 100 ns trajectory served to generate a series of configurations representative of a Boltzmann distribution in the ground state.

4000 initial configurations were sampled from the the ground-state Boltzmann distribution and propagated from time $t = 0$ on the excited-state potential energy surface, as described in Chapter 1, Section 2. The difference between the potential energy surfaces of the ground and L_a states of the indole chromophore in tryptophan-28 of the lysozyme molecule was implemented via a difference in partial charges on the indole atoms, as shown in Figure 2.1 and Table 2.1.[14,27–29,37,38] The Stokes shift was then estimated to be the energy difference, averaged over initial configurations, between the excited and ground states as a function of time as previously derived in Equation (1.6). To decompose the Stokes shift into group-specific contributions, as described in Equation (1.11), an indirect approach was taken. This approach will be explained in detail in the following section.

The simulations above were repeated for two subsets of frozen simulations. In this context, "frozen" refers to coordinate fixation at the moment of photo-excitation. In the first subset, deemed "frozen protein" simulations, the coordinates of the protein and the indole chromophore were fixed. In the second subset, "frozen protein and ions", the coordinates of the protein, the indole chromophore, and the ions were fixed. More information regarding the simulation parameters can be found in the Supplementary Materials section in Appendix A.

Atom	Ground	L_a	Difference
NE1	-0.10	0.07	0.17
HE1	0.31	0.31	
CD1	-0.14	-0.09	0.05
HD1	0.14	0.14	
CD2	0.00	-0.07	-0.07
CE3	0.14	-0.27	-0.41
HE3	0.14	0.14	
CG	-0.21	-0.06	0.15
CB	0.00	0.00	
CZ3	-0.14	-0.13	0.01
HZ3	0.14	0.14	
CH2	-0.14	-0.20	-0.06
HH2	0.14	0.14	
CZ2	-0.14	-0.27	-0.13
HZ2	0.14	0.14	
CE2	0.00	0.01	0.01

Table 2.1: Charges used on the indole group in the L_a state

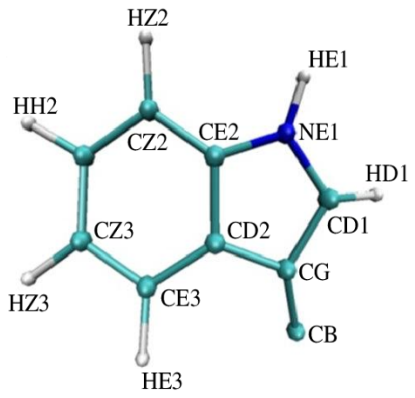

Figure 2.1: The indole chromophore of tryptophan-28 of lysozyme

2.2 Decomposition of Coulomb Interactions

To obtain the contributions from each of the components in Equation (1.9), the Coulomb energy, the only component of contributing to the difference between the ground and excited state potentials in our model, was decomposed based on group interactions. The original Coulomb potential is a sum of pair interactions. However, in the Ewald representation of the Coulomb energy, the reciprocal space portion of the Coulomb energy is not pairwise decomposable. Therefore, an indirect approach was taken. The indole-protein, indole-water, and indole-ion interactions were obtained by subtraction of the Coulomb reciprocal space contributions obtained by setting charges of individual groups (indole, protein, water, ion), or combinations of them, equal to zero. The individual groups are defined in Table 2.2.

Subscript	Definition
i	indole
p'	protein (excluding indole)
w	water
I	ions

Table 2.2: Definitions of symbols for charge groups

The potential energies were re-calculated using the previously simulated excited state trajectory and topology files with charges of different groups set to 0. This is outlined in the Table 2.3 and eqs. (2.1) to (2.4), in which V_{ij} denotes the interaction

between groups i and j. The indole-protein, indole-water, and indole-ion interactions were isolated via eqs. (2.5) to (2.7)

$$V_1(t) = V_{ii}(t) + V_{ip'}(t) + V_{p'p'}(t) \tag{2.1}$$

$$V_2(t) = V_{ii}(t) + V_{iI}(t) + V_{II}(t) \tag{2.2}$$

$$V_3(t) = V_{ii}(t) + V_{iw}(t) + V_{ww}(t) \tag{2.3}$$

$$V_4(t) = V_{ii}(t) \tag{2.4}$$

Charge Group Modification	Potential Energy
$q_w = q_I = 0$	$V_1(t)$
$q_w = q_{p'} = 0$	$V_2(t)$
$q_{p'} = q_I = 0$	$V_3(t)$
$q_w = q_{p'} = q_I = 0$	$V_4(t)$

Table 2.3: Scheme for recalculated potential energies based on modified charges

$$V_{ip'}(t) = V_1(t) - V_4(t) - V_{p'p'}(t) \tag{2.5}$$

$$V_{iw}(t) = V_3(t) - V_4(t) - V_{ww}(t) \tag{2.6}$$

$$V_{iI}(t) = V_2(t) - V_4(t) - V_{II}(t) \tag{2.7}$$

Because the Stokes shift involves the difference between the potential energy of the ground and excited states, the ground and excited modified potentials were used to calculate the protein, water, and ion contributions, as shown in eqs. (2.8) to (2.10).

$$\Delta E_{protein}(t) = V_{ip}^e(t) - V_{ip}^g(t) = V_1^e(t) - V_1^g(t) - V_4^e(t) + V_4^g(t) - V_{p'p'}^e(t) + V_{p'p'}^g(t) \tag{2.8}$$

$$\Delta E_{water}(t) = V_{iw}^e(t) - V_{iw}^g(t) = V_3^e(t) - V_3^g(t) - V_4^e(t) + V_4^g(t) - V_{ww}^e(t) + V_{ww}^g(t) \tag{2.9}$$

$$\Delta E_{ions}(t) = V_{iI}^e(t) - V_{iI}^g(t) = V_2^e(t) - V_2^g(t) - V_4^e(t) + V_4^g(t) - V_{II}^e(t) + V_{II}^g(t) \tag{2.10}$$

Considering that the water-water, protein-protein, and ion-ion interactions are the same between the ground and excited states, eqs. (2.8) to (2.10) become the final equation scheme as shown in eqs. (2.11) to (2.13).

$$S_{protein}(t) = V_1^e(t) - V_1^g(t) - V_4^e(t) + V_4^g(t) \tag{2.11}$$

$$S_{water}(t) = V_3^e(t) - V_3^g(t) - V_4^e(t) + V_4^g(t) \tag{2.12}$$

$$S_{ions}(t) = V_2^e(t) - V_2^g(t) - V_4^e(t) + V_4^g(t) \tag{2.13}$$

Chapter 3

Results and Discussion

3.1 Viability of Initial System Configuration

To test whether we had thoroughly sampled the lysozyme ground state, we tracked quantities like the ground state protein-indole interaction energy. The concern is that slow processes in the ground state, such as loop-hopping between two states, would not be sampled well in a 100 nanosecond trajectory. The ground state protein-indole interaction energy as a function of time can be found in Figure 3.1. It shows no evidence of long-timescale processes in the ground state.

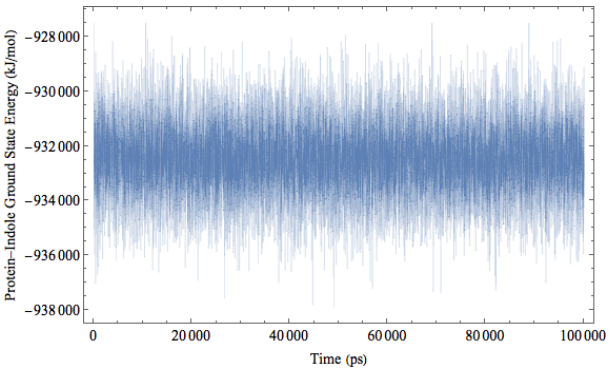

Figure 3.1: Ground state protein-indole interaction energy as a function of time in a 1 nanosecond ground state simulation

The distribution of the protein-indole ground state energy is shown in Figure 3.2. The protein-indole ground state energy is normally distributed, again suggesting that there are no long-timescale events occurring that would skew the ground state energy distribution, thus affecting the ground state configuration distribution from which the configurations at $t = 0$ were sampled. Prominently, the lack of a bimodal, or multimodal, distribution in Figure 3.2 confirms that no slow processes are occurring because these would generally result in two (or more) minimum free energy configurations.

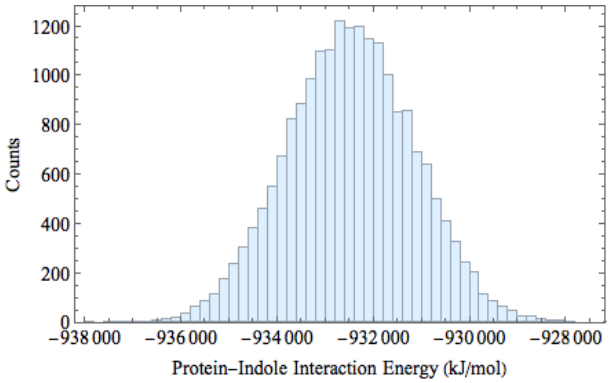

Figure 3.2: Ground state protein-indole interaction energy distribution

3.2 Flexible System

The Stokes shift was calculated for the flexible system using the construction in Equation (1.11), as displayed in Figure 3.3. The indole-ion interactions, like

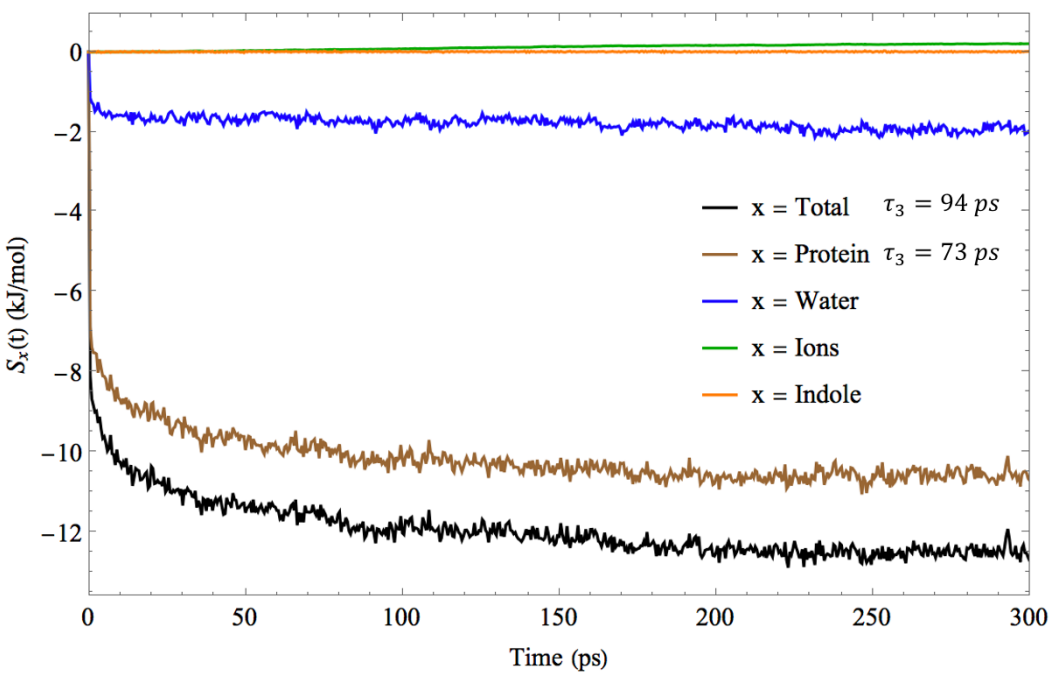

Figure 3.3: Stokes shift for the flexible system

the indole-indole interactions, are negligible in comparison with the water-indole and protein-indole terms. Thus, $S_{water}(t)$ and $S_{protein}(t)$ constitute the dominant contributions to the total Stokes shift, $S_{total}(t)$.

Figure 3.3 provides qualitative evidence of the inconsistencies within Halle and Nilsson's dielectric continuum model for the Stokes shift. Their dielectric continuum model predicts that the water and protein contributions to the Stokes shift will be equal and opposite.[14,31] In our flexible system simulations, the protein component shows a long-time decay that is far different from the mirror image of that of water.

The curves, $S_x(t)$ were fitted using additive exponential decays of the form of Equation (3.1). Each of the three exponential components were restricted to time regions corresponding to the three components of the Stokes shift. The first exponential component, with time constant τ_1, corresponds to the inertial decay region and is therefore restricted between 0 and 1 picoseconds. As mentioned previously in Chapter 1, Section 2, a Gaussian fit is used to represent the inertial decay region. The second component, corresponding to hindered water rotations, was bounded by 5 and 20 picoseconds. The third and final region, corresponding to the slow component, ranged from 20 to 300 picoseconds.

$$S_x(t) = -S_{x,\infty}\left(1 - (1 - c_2 - c_3)e^{-\left(\frac{t}{\tau_1}\right)^2} - c_2 e^{-\left(\frac{t}{\tau_2}\right)} - c_3 e^{-\left(\frac{t}{\tau_3}\right)}\right) \tag{3.1}$$

It is important to note that the form of Equation (3.1) is somewhat arbitrary, and the fits should be regarded as a heuristic tool. In some cases, we know that, on empirical grounds, there are more than three decay components contributing to the Stokes shift.

S_x (kJ/mol)	$S_{x,\infty}$ (kJ/mol)	τ_1 (ps)	c_2	τ_2 (ps)	c_3	τ_3 (ps)
Total	12.65	0.2905	0.1644	6.861	0.1753	94.0299
Protein	10.65	0.2942	0.1553	7.268	0.1750	73.2924
Water	2.429	0.2936	0.0922	3.202	0.5670	1201.04

Table 3.1: Exponential fit to Stokes shift data

The time constant for the third region, τ_3, is similar for protein contribution and the total Stokes shift, while that of water is vastly different, suggesting that the protein contribution constitutes the majority long-term component of the Stokes shift. This striking congruence is confirmed by the normalized Stokes shifts, $S_3'(t)$, displayed in Figure 3.4. The normalized Stokes shifts, though very similar, are not exactly equal, showing that the total Stokes shift has a component other than the protein. The fitted parameters, summarized in Table 3.1, reveal several features of the relaxation process following photo-excitation. The time constant of the slow component of protein, $\tau_{3,protein}$, is two orders of magnitude smaller than that of the slow component of water, $\tau_{3,water}$, reinforcing, as mentioned in the discussion of Figure 3.3, that the protein and water contributions do not mirror each other. Past the inertial drop, the water contribution to the Stokes shift appears to be dominated by the long-timescale component, as can be seen by the weight of the slow component, $c_{3,water} = 0.5670$. In Figure 3.3, it appears that the ionic contribution to the Stokes shift, not the contribution from protein as predicted by Halle and Nilsson,

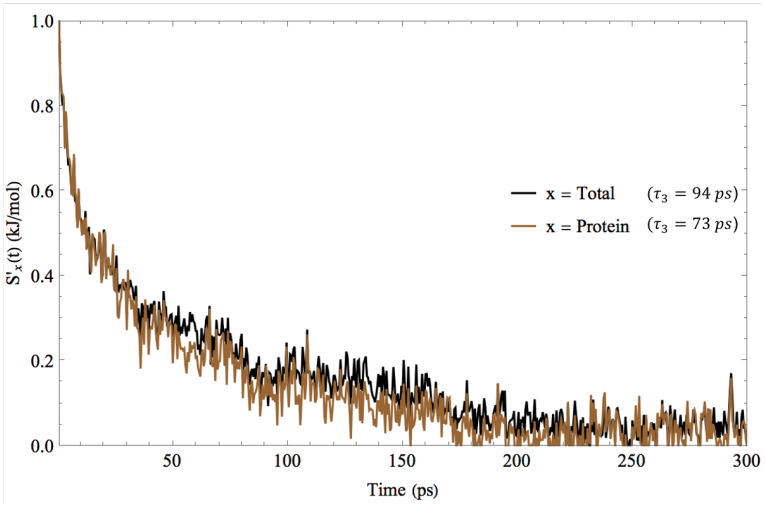

Figure 3.4: The normalized Stokes shift of the flexible system

is equal and opposite to the contribution from water. To test this observation, the indole-ionic and indole-water Stokes shifts in the third region, ranging from 20-300 picoseconds, were fitted with a linear model, as shown in Equation (3.2).

$$S_x(t) = kt + b \tag{3.2}$$

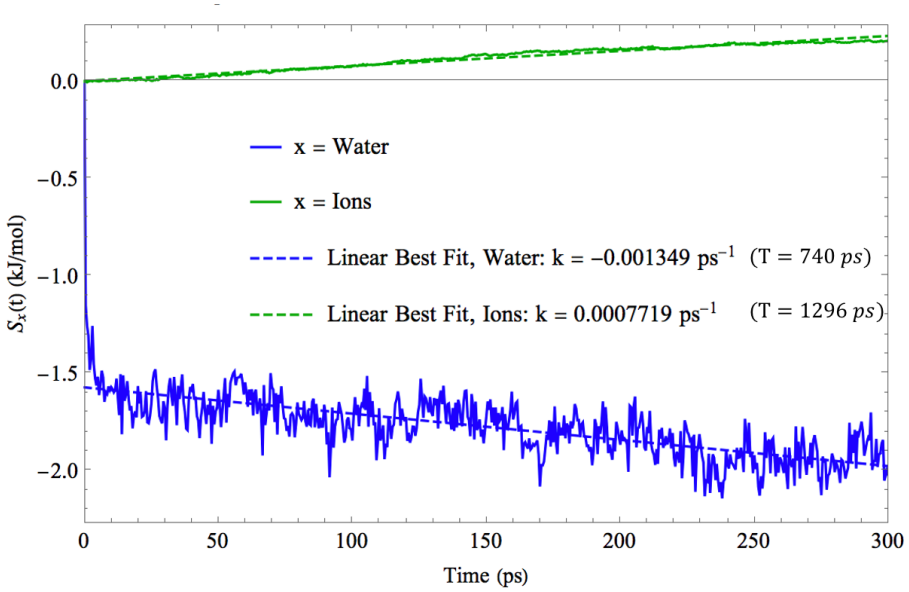

Figure 3.5: When the long-timescale ranges of the water and ionic Stokes shifts are fit with the linear model in Equation (3.2), the slopes are not equal in magnitude and opposite.

The intercept, b, is arbitrary, but the slope, k, is indicative of the relationship between the water and ionic responses. Defining a constant T, with units of picoseconds, a meaningful long-timescale correlation can be extracted from the water and

ionic Stokes shifts, as shown in Equation (3.3).

$$T = \left|\frac{1}{k}\right| (ps) \tag{3.3}$$

The linear fits of the water and ionic Stokes shifts, as shown in Figure 3.5, yield similar, although not equal, and opposite responses. As we will demonstrate below, most of the water response reflects coupling to ionic motion, accounting for the similar T values for water and ions. However, another part of the water response arises from the type of protein-water motion described by Li et al. 2007. This relationship will be further discussed in Section 3 through simulations in which protein motion is frozen.

3.3 Frozen Protein System

The Stokes shift simulations and calculations were repeated for the frozen protein system, shown in Figure 3.6. $S_{water}(t)$ was fitted using Equation (3.1). However, the weight of the long-timescale component, c_3, was negligible for $S_{total}(t)$. $S_{total}(t)$ was then re-fitted using Equation (3.4). The parameters for both fits are displayed in Table 3.2.

$$S_x(t) = -S_{x,\infty}\left(1 - (1 - c_2)e^{-\left(\frac{t}{\tau_1}\right)^2} - c_2 e^{-\left(\frac{t}{\tau_2}\right)}\right) \tag{3.4}$$

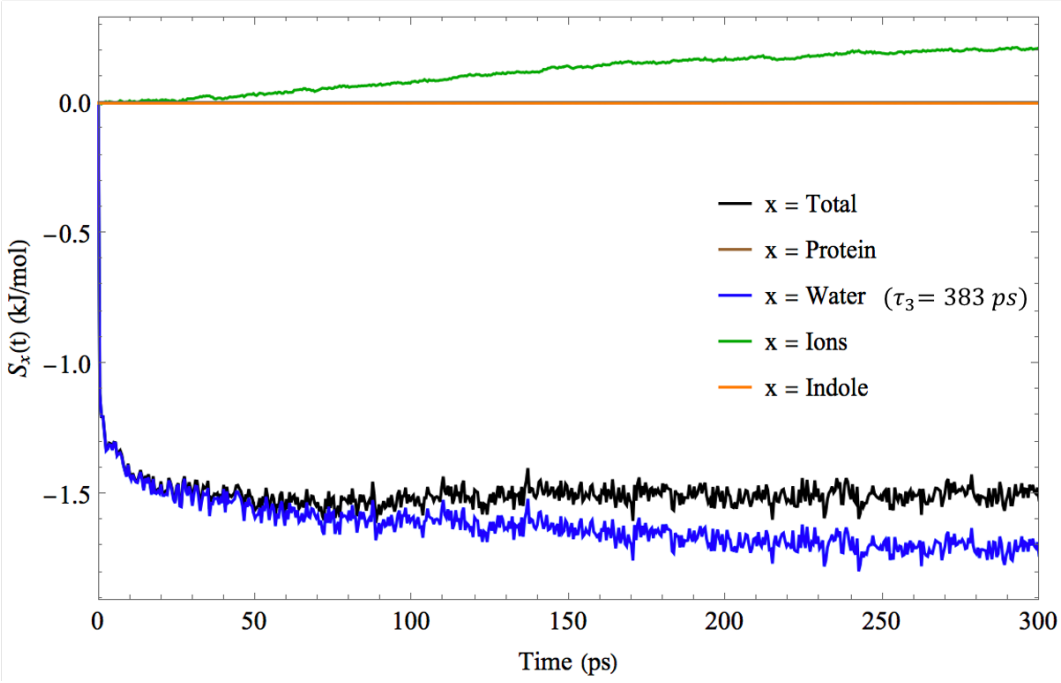

Figure 3.6: Stokes shift for the frozen protein system

S_x (kJ/mol)	$S_{x,\infty}$ (kJ/mol)	τ_1 (ps)	c_2	τ_2 (ps)	c_3	τ_3 (ps)
Total	1.497	0.2978	0.3426	10.58	——	——
Water	1.883	0.2925	0.1864	7.646	0.2005	383.911

Table 3.2: Exponential fit to frozen protein Stokes shift data

Three important trends emerge from Figure 3.6. The first of these is that the total Stokes shift of the frozen protein system is an order of magnitude less than the total Stokes shift of the flexible system in Figure 3.3. The order of magnitude difference of the system relaxation upon fixing the protein coordinates at photoexcitation, in comparison with that of the flexible system, suggests that protein comprises the majority of the chromophore's surroundings that participate in relaxation following photo-excitation. Thus, the chromophore is indeed "buried" within the protein, as has been previously noted by the Bagchi group.

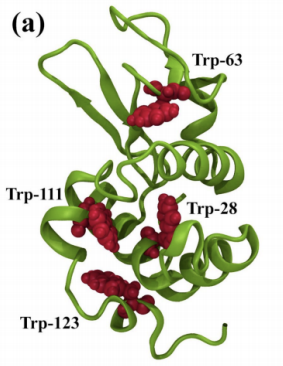

Figure 3.7: Trp-28 in is "buried" in lysozyme, taken from Bagchi et al. 2017

The ionic contribution to the Stokes shift, $\|S_{ions}(t)\| \sim 0.2$ kJ/mol, is negligible in the flexible system, with $\|S_{total}(t)\| \approx 10$ kJ/mol, but significant in the frozen system, with $\|S_{total}(t)\| = 2$ kJ/mol. This leads to the second key observation from the Stokes shift plot in Figure 3.2, that the slow component of $S_{ions}(t)$ appears to now exactly mirror the the slow component of $S_{water}(t)$. The apparent relationship was verified by fitting the water and ionic Stokes shifts using the linear model in Equation (3.2). As shown in Figure 3.8, the water and ionic Stokes shifts have equal and opposite slopes when protein is frozen, indicating that water and ionic responses are intimately coupled. This relationship will be analyzed under an additional constraint in Section 4, where both the protein and the ions are frozen.

Finally, the third observation from Figure 3.6 and Table 3.2 is that all time-dependence of $S_{total}(t)$ past ~ 50 picoseconds disappears when the protein is frozen at the moment of photo-excitation. This is qualitatively seen in the Stokes shift because the total Stokes shift rapidly converges to a constant. The Stokes fit pa-

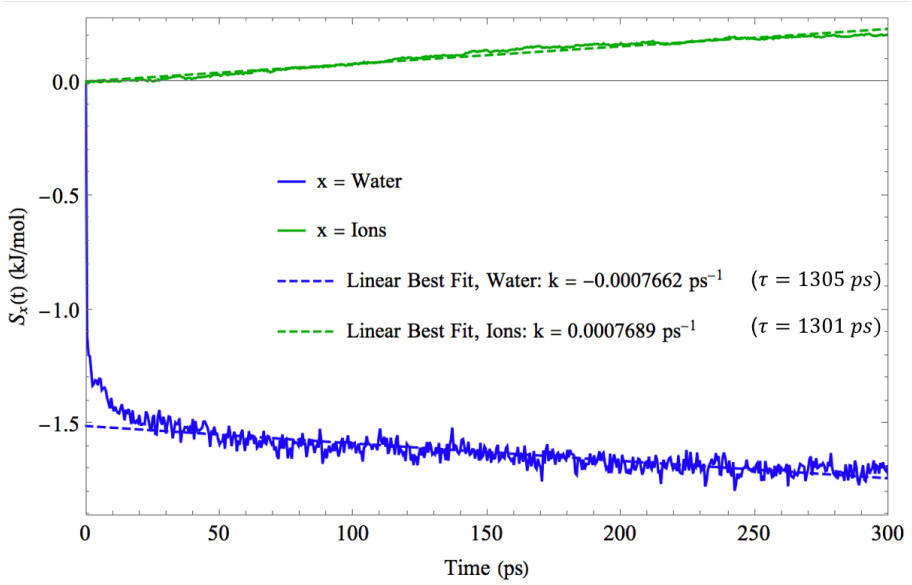

Figure 3.8: When the long-timescale ranges of the water and ionic Stokes shifts in the frozen protein system are fit with the linear model in Equation (3.2), the slopes are equal in magnitude an opposite.

rameters in Table 3.2 numerically confirm this observation. As presented in Equation (3.4), the best fit for the the total Stokes shift, $S_{total}(t)$, neglects the third decay component. When we include the $c_3 e^{-\left(\frac{t}{\tau_3}\right)}$ term, the weight of the third component, $c_{3,total}$, is an order of magnitude smaller than that of the flexible system, further suggesting the slow component of the total Stokes shift has disappeared upon freezing. We propose that the water and ionic Stokes shifts nearly, if not completely, cancel each other, thus allowing the total Stokes shift to rapidly converge to a constant when protein motion is frozen. The physical explanation for this is discussed in Section 5.

3.4 Frozen Protein and Ions System

In response to the newly observed ionic contribution to the Stokes shift in the frozen protein system, the simulation method was repeated for a system in which both the protein and the ions were frozen at the moment of photo-excitation. We average excited state trajectories over initial protein and ion distributions representative of a thermal distribution because the initial configurations were sampled from the ground state trajectory in which the ions move, only restricting ion and protein motion following photo-excitation. The Stokes shifts generated from this system are presented in Figure 3.9.

Figure 3.9 confirms that the slow component of the total Stokes shift, which now only contains a water component, disappears when both the proteins and the

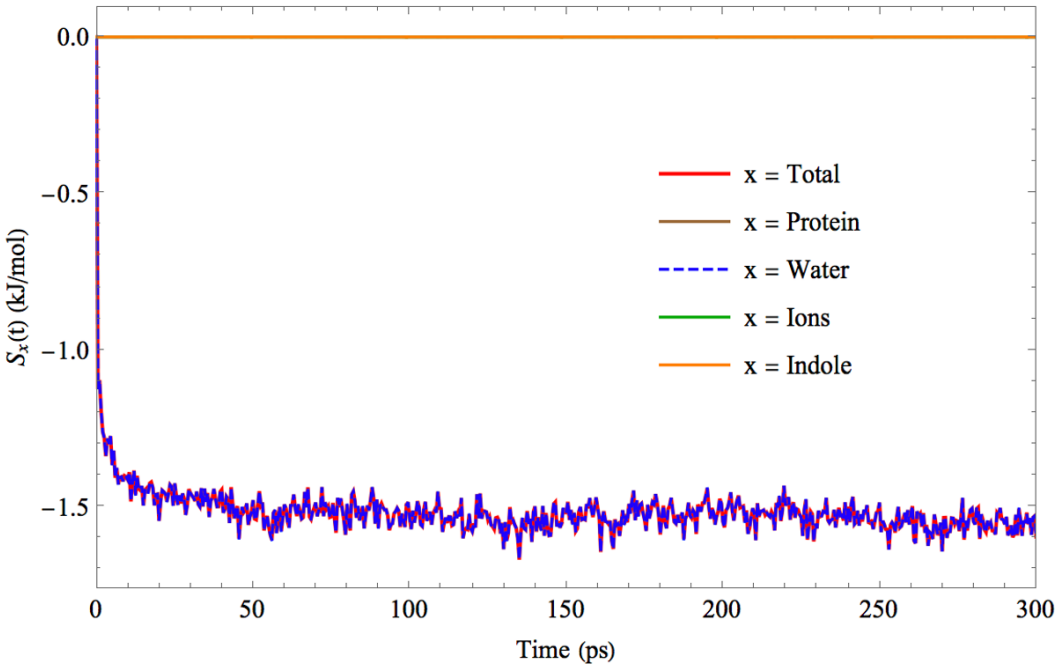

Figure 3.9: The Stokes shift of the frozen protein and ions system

ions are frozen. $S_{total}(t)$ and $S_{water}(t)$ were fit using Equation (3.1). Exhibited by the rapid convergence to a constant in both $S_{total}(t)$ and $S_{water}(t)$, the weight of the long-timescale component, c_3, was negligible for each fit and, as such, both data sets were re-fitted with the modified model in Equation (3.4), with parameters displayed in Table 3.3. The fits to $S_{total}(t)$ and $S_{water}(t)$ are nearly exact. When included in the fit, the weight of the long-timescale component, c_3, was nearly zero in both fits, providing further evidence that the long-timescale component disapppears for $S_{total}(t)$ and $S_{water}(t)$ when the motions of the protein and the ions are fixed at photo-excitation.

S_x (kJ/mol)	S_∞ (kJ/mol)	τ_1 (ps)	c_2	τ_2 (ps)	c_3	τ_3 (ps)
Total	1.551	0.2024	0.3877	3.432	——	——
Water	1.552	0.2023	0.3884	3.439	——	——

Table 3.3: Exponential fit to frozen protein and ions Stokes shift data

Deserving special attention is the trend between the water and ions in the two frozen systems. The aforementioned equal and opposite long-timescale components of the ionic and water contributions in the frozen protein system are not seen in the frozen protein and ions system. The Stokes shifts were again fit with the linear model in Equation (3.2), with results displayed in Figure 3.10. In our system, with less than ten counter-ions, we find that the ionic contribution to the total energy difference comprising the Stokes shift is not, as remarked upon by Bagchi et al. 2017, "negligible compared to others." Instead, we find that the ions are significant in the

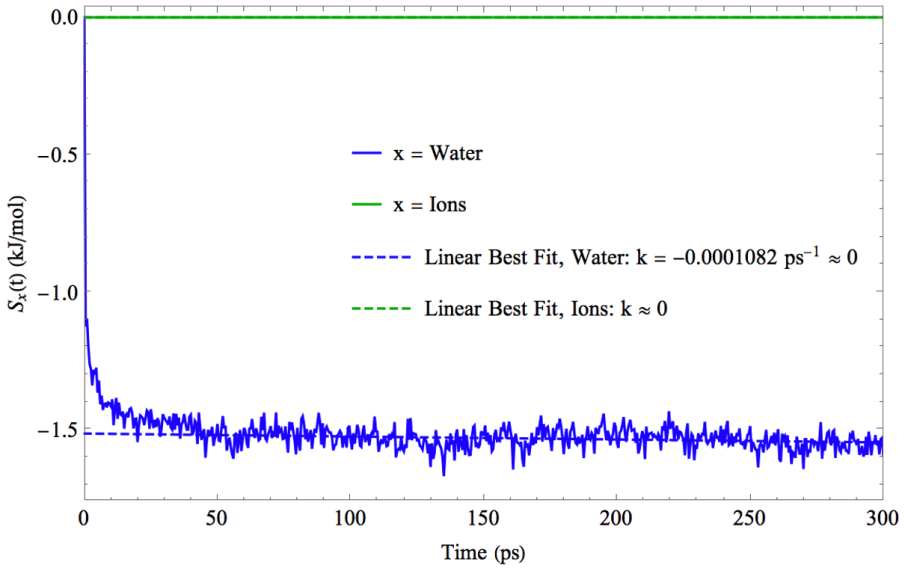

Figure 3.10: When the long-timescale regions of the water and ionic Stokes shifts in the frozen protein an ions system are fit with the linear model in Equation (3.2), the slopes of both linear fits are 0.

Stokes shift of tryptophan-28 of lysozyme due to the small magnitude of the total Stokes shift resulting from the buried location of the residue. Most significantly, we find that the slow component of water disappears when ions are frozen along with the protein.

3.5 Discussion

In response to our main motivation, the results of this work directly contradict the faulty analysis by Bagchi et al. which proposed that the long-timescale dynamics in the Stokes shift of Trp-28 of lysozyme is evidence of biological water near the protein. The Stokes shift of the flexible system (Figure 3.3) reiterates the previous finding from Li et al. 2007 that the total Stokes shift includes water and protein contributions due to protein relaxation and protein-coupled response from the water. In simulations where protein motion is frozen, the water contribution to the Stokes shift appears to have long-timescale dynamics, which had been calculated by Bagchi et al. and interpreted as "biological water." However, when ions are frozen as well, the slow component in the water response disappears, proving that the slow component of water with only protein frozen reflects slow ion dynamics, not intrinsically slow water.

We propose that this slow component in the water contribution represents a response to the slow ionic diffusion far away from the protein. Polarization charge, depicted in Figure 3.11, is the layer of counter-charge within water near an ion that

accumulates when the relatively positive hydrogen atoms of the water molecules reorient around the negatively charged ion to stabilize the charge. As the ions far away from the protein diffuse over time, the always-present layer of polarization charge also diffuses, giving rise to an indole-water contribution that is equal in magnitude and opposite to the indole-ion contribution.

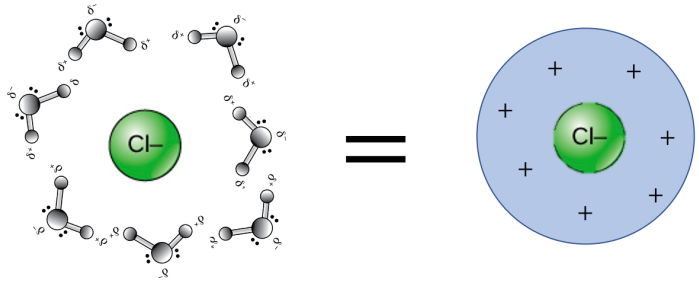

Figure 3.11: Polarization charge is the layer of counter-charge within water near an ion

In the fully flexible system, we propose that water contributes to the Stokes shift through two indirect ways: coupled protein-water motion and coupled ion-water motion. In the frozen protein system, where there is no protein relaxation to induce coupled protein-water motion, the water and ionic components of the Stokes shift are nearly, if not exactly, equal and opposite (Figure 3.8).

Halle and Nilsson's dielectric continuum model, which, as they originally proposed, is a poor model for describing protein-water coupling, turns out to explain the near cancellation between the water and ionic contributions with protein frozen. Under our dielectric continuum model, the total difference between the ground and excited state interaction energies, taken as a function of ion coordinates, $\Delta E_{total}(x_{ions})$, consists of water and ionic contributions, as shown in Equation (3.5). The ionic contribution, $\Delta E_{ions}(x_{ions})$, is represented as the gas-phase Coulomb interaction potential between the ions and the indole chromophore, $\Delta E^{o}_{total}(x_{ions})$.

$$\Delta E_{total}(x_{ions}) = \Delta E_{water}(x_{ions}) + \Delta E_{ions}(x_{ions}) \tag{3.5}$$

In water, modeled as a continuum with dielectric constant ϵ, $\Delta E_{total}(x_{ions})$ can be calculated from the gas phase indole-ionic interaction potential because the ions are far from the indole chromophore, as shown in Equation (3.6).

$$\Delta E_{total}(x_{ions}) = \frac{1}{\epsilon}\Delta E^{o}_{ions}(x_{ions}) \tag{3.6}$$

Combining Equations (3.5) and (3.6), we solve for the long-timescale water contri-

bution to $\Delta E_{total}(x_{ions})$ in Equation (3.7).

$$\Delta E_{water}(x_{ions}) = -\left(1 - \frac{1}{\epsilon}\right) \Delta E^{o}_{ions}(x_{ions}) \tag{3.7}$$

As mentioned in Chapter 1, Section 4, $\epsilon \gg 1$ with water as a solvent, so the water contribution to the interaction between chromophore and surroundings should be approximately equal and opposite to the ionic contribution, as shown in Equation (3.8). Thus, Equation (1.20) from the Halle and Nilsson view should actually be Equation (3.9).

$$\Delta E_{water}(x_{ions}) \approx -\Delta E^{o}_{ions}(x_{ions}) \tag{3.8}$$

$$S_{water}(t) = -\left(1 - \frac{1}{\epsilon}\right) S_{ions}(t) \tag{3.9}$$

The long-timescale equal and opposite trends of the water and ionic Stokes shifts in Figure 3.8 validate our model in Equation (3.9). The model is further validated by the disappearance of the long-timescale water component when both protein and ions are frozen (Figure 3.10). Therefore, we conclude that the long-timescale component of the water contribution to the Stokes shift is a result of coupled protein-water and ion-water motion.

Chapter 4

Conclusion

Understanding of the photophysical behavior underlying tryptophan fluorescence is a requisite for its use of in the study of protein dynamics. The difference between the emission and excitation energy of a photon as the tryptophan residue's indole chromophore fluoresces over time, otherwise known as the time-dependent fluorescence Stokes shift, is highly reflective of the systematic relaxation of the chromophore's surrounding environment. The long-timescale dynamics of the Stokes shift have been explained using the "biological water" model, wherein it has been proposed that the slow dynamics are a product of the rapid exchange of "bound" water molecules, constituting an ice-like hydration in the immediate vicinity of a protein, with freely-moving water molecules far away.[17–21,23] As an alternative, previous studies from our group, as well as others, have explained that the long-timescale component of the total Stokes shift is reflective of protein dynamics as the protein relaxes to the chromophore's changing charge distribution, as well as protein-coupled water motion. However, the slow component of the total Stokes shift is not reflective of the intrinsic water dynamics, as proposed by the "biological water" model.[14,27–31] Linear response theory and non-equilibrium methods have been used to simulate the systematic relaxation following photo-excitation of the chromophore as part of the investigation of the origin of the long-timescale dynamics of the Stokes shift.

Previous studies from our group proposed a non-equilibrium simulation technique wherein the protein is immobilized at the moment of photo-excitation.[27–29] By rendering the protein unable to relax in response to the changing charge distribution of the chromophore, our group showed that the long-timescale component of the Stokes shift disappeared which undermined the attribution of slow dynamics to the water molecules. Bagchi and co-workers, proponents of "biological water", utilized this technique with tryptophan-28 of lysozyme and proposed that slow dynamics remain even when the protein is frozen, therefore substantiating the attribution of

slow dynamics to water as part of the "biological water" model. With this claim as the main motivation for this work, we have conducted an investigation into the time-dependent fluorescence Stokes shift of indole in Trp-28 of lysozyme and we propose that the long-timescale component of the Stokes shift is a result of protein motion, coupled protein-water motion, and coupled ion-water motion.

Using initial configurations sampled from a Boltzmann distribution in the ground state, we propagated the system from time $t = 0$ on the excited-state potential energy surface and calculated the total Stokes shift of the system, $S_{total}(t)$, as the time-dependent energy difference between the ground and excited states. The energy difference, which consists of changes in the Coulombic interactions between the indole chromophore and its environment, was decomposed to calculate the protein, water, and ionic contributions to the Stokes shift. The resulting Stokes shifts were fitted using a sum of three exponentials representing each of the expected components to the Stokes shift: the inertial decay region between 0 and 1 picoseconds, the region between 5 and 20 picoseconds dominated by water librations, and the long-timescale region between 20 and 300 picoseconds.

For the flexible system, where all components of the environment are capable of responding, our results in Figure 3.3 show that the total, protein, water, and ionic Stokes shifts have long-timescale components. We observe a mirrored trend between the long-timescale components of the water and ionic Stokes shifts. When fit to a linear model, the long-timescale components prove to be similar, though not equal, and opposite responses.

As previously described by Li et al. 2007, we attribute the aforementioned incongruity to coupled protein-water motion that arises when the protein is able to relax in response to the changing charge distribution of the chromophore. We tested our theory through the Singer group's frozen protein technique, with results displayed in Figure 3.6. We find that the slow component of the total Stokes shift disappears with protein frozen. However, we find that a slow component remains in the water and ionic contributions. Fitting the long-timescale regions of the water and ionic Stokes shifts to a linear model, we find that the slopes of the curves (with units of picoseconds^{-1}) are equal in magnitude and opposite (Figure 3.8). Furthermore, the slopes of the linear fits to the water and ionic curves were effectively 0 in subsequent simulations with both the protein and the ions frozen, as shown in Figure 3.10.

Combining the results of the flexible, frozen protein, and frozen protein and ions systems, we conclude that, due to neglect of ions, Bagchi and co-workers erroneously interpreted the long-timescale component in the water response for the frozen pro-

tein system as evidence of "biological water". Alternatively, we propose that the persistence of the slow component of the water response between the flexible and frozen protein systems results from the motion of water within the polarization layer of counter-charge near an ion as the ion slowly diffuses, far away from the chromophore. We emphasize the linear fits to the long-timescale components of the water and ionic contributions as the most significant evidence behind our conclusion (Figures 3.5, 3.8, and 3.10).

Addressing the secondary motivation of this work, we find no support for the Halle and Nilsson model that calculates that the protein contribution should mirror that of water. Instead, we present a new dielectric continuum model which calculates that the long-timescale component of the water contribution to the Stokes shift should be equal in magnitude and opposite to that of the ionic contribution. The clearly equal and opposite trends of water and ionic Stokes shifts in the frozen protein system, as well as the disappearance of the long-timescale component of the water curve in the frozen protein and ions system, confirm the validity of our model. In closing, this work shows that the long-timescale component of the water contribution to the Stokes shift of tryptophan-28 in lysozyme results from, as suggested previously, coupled protein-water motion as well as our novel finding of coupled ion-water motion, therefore nullifying the claim that the long-timescale component is due to intrinsically slow dynamics of water as part of the "biological water" model.

Appendix A

Supplementary Materials

A.1 Simulation Parameters

Lysozyme (PDB:1AKI)

A periodically-replicated cubic box of side length 80Å was used. The GROMOS 54A7 force field was used.[39] Previous calculations[37] were used to implement the difference in partial atomic charges of indole between the ground and excited states. Long range electrostatics were handled using the smooth Particle Mesh Ewald (SPME) algorithm[40,41] with a real-space cutoff radius of 10Å. Long range dispersion corrections were applied for both energy and pressure. The cutoff length for the Lennard-Jones potentials was set to 10Å. For the ground state simulation, a canonical velocity-rescaling thermostat[42] and a 0.1 picosecond time constant were used to maintain the system temperature at 298 K. Additionally, isotropic Berendsen pressure coupling[43] was used with a 0.5 picosecond time constant to maintain the pressure at 1.0 bar in all directions.

Frozen Protein and Frozen Protein & Ions Simulations

For both the frozen protein and the frozen protein & ions simulations, the parameters were the same as above, with two modifications. The freeze groups were specified as protein or protein & ions. Accordingly, pressure coupling was disabled.

Bibliography

(*1*) Chen, Y., and Barkley, M. D. (1998). Toward Understanding Tryptophan Fluorescence in Proteins. *Biochem. J. 37*, 9976–9982.

(*2*) Ghisaidoobe, A. B., and Chung, S. J. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy transfer techniques. *Int J of Mol Sci 15*, 22518–22538.

(*3*) Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T.-M., Rech, I., Cova, S., Xun, L., and Xie, X. S. (2003). Protein Conformational Dynamics Probed by Single-Molecule Electron Transfer. *Science 302*, 262–266.

(*4*) Levitt, M., and Sharon, R. (1988). Accurate Simulation of Protein Dynamics in Solution. *Proc. Nat. Acad. Sci. 85*, 7557–7561.

(*5*) Otting, G., Liepinsh, E., and Wuthrich, K. (1991). Protein Hydration in Aqueous Solution. *Science 254*, 974–980.

(*6*) Teeter, M. M. (1991). Water-Protein Interactions: Theory and Experiment. *Annu. Rev. Biophys. Biophys. Chem. 20*, PMID: 1867726, 577–600.

(*7*) Daniel, R. M., Finney, J. L., Stoneham, M., and Halle, B. (2004). Protein Hydration Dynamics in Solution: A Critical Survey. *Philos. Trans. R. Soc. B 359*, 1207–1224.

(*8*) Levy, Y., and Onuchic, J. N. (2006). Water Meditation in Protein Folding and Molecular Recognition. *Annual Review of Biophysics and Biomolecular Structure 35*, PMID: 16689642, 389–415.

(*9*) Perera, L., and Berkowitz, M. L. (1992). Ultrafast Solvation Dynamics in a Stockmayer Fluid. *J. Chem. Phys. 97*, 5253–5254.

(*10*) Carter, E. A., and Hynes, J. T. (1991). Solvation Dynamics for an Ion Pair in a Polar Solvent: Time-Dependent Fluorescence and Photochemical Charge Transfer. *J. Chem. Phys. 94*, 5961.

(*11*) Rosenthal, S. J., Xie, X., Du, M., and Fleming, G. R. (1991). Femtosecond Solvation Dynamics in Acetonitrile: Observation of the Inertial Contribution to the Solvent Response. *J. Chem. Phys. 95*, 4715–4718.

(*12*) Hsu, C.-P., Song, X., and Marcus, R. A. (1997). Time-Dependent Stokes Shift and Its Calculation from Solvent Dielectric Dispersion Data. *J. Phys. Chem. B 101*, 2546–2551.

(*13*) Maroncelli, M., and Fleming, G. R. (1988). Computer Simulation of the Dynamics of Aqueous Solvation. *J. Chem. Phys. 89*, 5044.

(*14*) Nilsson, L., and Halle, B. (2005). Molecular Origin of Time-Dependent Fluorescence Shifts in Proteins. *Proc. Natl. Acad. Sci. U.S.A. 102*, 13867–13872.

(*15*) Kuntz, I., and Kauzmann, W. In, Anfinsen, C., Edsall, J. T., and Richards, F. M., Eds.; Adv. Protein Chem. Vol. 28; Academic Press: 1974, pp 239–345.

(*16*) Pethig, R. (1992). Protein-Water Interactions Determined by Dielectric Methods. *Annu. Rev. Phys. Chem. 43*, PMID: 1463572, 177–205.

(*17*) Pal, S. K., and Bagchi, B. (1997). Dielectric Relaxation of Biological Water. *J. Phys. Chem. B 101*, 10954–10961.

(*18*) Pal, S. K., Peon, J., and Zewail, A. H. (2002). Biological Water at the Protein Surface: Dynamical Solvation Probed Directly with Femtosecond Resolution. *Proc. Nat. Acad. Sci. 99*, 1763–1768.

(*19*) Pal, S. K., Peon, J., Bagchi, B., and Zewail, A. H. (2002). Biological Water: Femtosecond Dynamics of Macromolecular Hydration. *J. Phys. Chem. B. 106*, 12376–12395.

(*20*) Pal, S. K., Peon, J., and Zewail, A. H. (2002). Ultrafast Surface Hydration Dynamics and Expression of Protein Functionality: α-Chymotrypsin. *Proc. Nat. Acad. Sci. 99*, 15297–15302.

(*21*) Peon, J., Pal, S. K., and Zewail, A. H. (2002). Hydration at the Surface of the Protein Monellin: Dynamics with Femtosecond Resolution. *Proc. Nat. Acad. Sci. 99*, 10964–10969.

(*22*) Jimenez, R., Fleming, G. R., Kumar, P. V., and Maroncelli, M. (1994). Femtosecond Solvation Dynamics of Water. *Nature 369*, 471–473.

(*23*) Mondal, S., Mukherjee, S., and Bagchi, B. (2017). Decomposition of Total Solvation Energy Into Core, Side-Chains and Water Contributions: Role of Cross Correlations and Protein Conformational Fluctuations in Dynamics of Hydration Layer. *Chem. Phys. Lett. 683*, Ahmed Zewail (1946-2016) Commemoration Issue of Chemical Physics Letters, 29–37.

(*24*) Furse, K. E., and Corcelli, S. A. (2010). Molecular Dynamics Simulations of DNA Solvation Dynamics. *J. Phys. Chem. Lett 1*, 1813–1820.

(*25*) Abbyad, P., Shi, X., Childs, W., McAnaney, T. B., Cohen, B. E., and Boxer, S. G. (2007). Measurement of Solvation Responses at Multiple Sites in a Globular Protein. *J. Phys. Chem. B 111*, PMID: 17592867, 8269–8276.

(*26*) Golosov, A. A., and Karplus, M. (2007). Probing Polar Solvation Dynamics in Proteins: A Molecular Dynamics Simulation Analysis. *J. Phys. Chem. B 111*, PMID: 17249715, 1482–1490.

(*27*) Li, T., Hassanali, A. A., Kao, Y.-T., Zhong, D., and Singer, S. J. (2007). Hydration Dynamics and Time Scales of Coupled Water-Protein Fluctuations. *J. Am. Chem. Soc. 129*, PMID: 17319669, 3376–3382.

(*28*) Hassanali, A. A., Li, T., Zhong, D., and Singer, S. J. (2006). A Molecular Dynamics Study of Lys-Trp-Lys: Structure and Dynamics in Solution Following Photoexcitation. *J. Phys. Chem. B. 110*, PMID: 16722759, 10497–10508.

(*29*) Li, T., Hassanali, A. A., and Singer, S. J. (2008). Origin of Slow Relaxation Following Photoexcitation of W7 in Myoglobin and the Dynamics of Its Hydration Layer. *J. Phys. Chem. B 112*, PMID: 18729509, 16121–16134.

(*30*) Furse, K. E., and Corcelli, S. A. (2008). The Dynamics of Water at DNA Interfaces: Computational Studies of Hoechst 33258 Bound to DNA. *J. Am. Chem. Soc. 130*, PMID: 18767841, 13103–13109.

(*31*) Halle, B., and Nilsson, L. (2009). Does the Dynamic Stokes Shift Report on Slow Protein Hydration Dynamics? *J. Phys. Chem. B 113*, PMID: 19462949, 8210–8213.

(*32*) Laage, D., Elsaesser, T., and Hynes, J. T. (2017). Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev. 117*, PMID: 28248491, 10694–10725.

(*33*) Jungwirth, P. (2015). Biological Water or Rather Water in Biology? *The Journal of Physical Chemistry Letters 6*, PMID: 26266717, 2449–2451.

(*34*) Lindahl, Abraham, Hess, and van der Spoel GROMACS 2020 Source code., version 2020, 2020.

(*35*) Carter, D., He, J., Ruble, J., and Wright, B. (1997). The Structure of the Orthorhombic Form of Hen Egg-White Lysozyme at 1.5 Angstroms Resolution.

(*36*) Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987). The Missing Term in Effective Pair Potentials. *J. Phys. Chem. 91*, 6269–6271.

(*37*) Sobolewski, A. L., and Domcke, W. (1999). Ab Initio Investigations on the Photophysics of Indole. *Chem. Phys. Lett 315*, 293–298.

(*38*) Vivian, J. T., and Callis, P. R. (2001). Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophys. J. 80*, 2093–2109.

(*39*) Schmid, N., Eichenberger, A., Choutko, A., Riniker, S., Winger, M., Mark, A., and Gunsteren, W. (2011). Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J. 40*, 843–856.

(*40*) Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys. 98*, 10089–10092.

(*41*) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A Smooth Particle Mesh Ewald Method. *J. Chem. Phys. 103*, 8577–8593.

(*42*) Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys. 126*, 014101.

(*43*) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular Dynamics With Coupling to an External Bath. *J. Chem. Phys. 81*, 3684–3690.